

OPEN

Antifreeze protein complements cryoprotective dehydration in the freeze-avoiding springtail *Megaphorura arctica*

Laurie A. Graham¹, Marie E. Boddington¹, Martin Holmstrup^{2,3} & Peter L. Davies^{1*}

The springtail, *Megaphorura arctica*, is freeze-avoiding and survives sub-zero temperatures by cryoprotective dehydration. At the onset of dehydration there is some supercooling of body fluids, and the danger of inoculative freezing, which would be lethal. To see if the springtails are protected by antifreeze proteins in this pre-equilibrium phase, we examined extracts from cold-acclimated *M. arctica* and recorded over 3 °C of freezing point depression. Proteins responsible for this antifreeze activity were isolated by ice affinity. They comprise isoforms ranging from 6.5 to 16.9 kDa, with an amino acid composition dominated by glycine (>35 mol%). Tryptic peptide sequences were used to identify the mRNA sequence coding for the smallest isoform. This antifreeze protein sequence has high similarity to one characterized in *Hypogastrura harveyi*, from a different springtail order. If these two antifreeze proteins are true homologs, we suggest their origin dates back to the Permian glaciations some 300 million years ago.

The springtail, *Megaphorura arctica* Tullberg 1876, is widely distributed in the northern parts of the Palaearctic region where it is common and abundant along sea shores¹. Aggregations of the species are often found in decaying seaweed of the supra-littoral zone, under stones and in moss below bird colonies^{2,3}. Measurements of the winter temperature of its habitat show that the species is able to survive temperatures of –20 °C or lower⁴. Freeze-tolerance (i.e. tolerance of organismal ice) has never been observed in springtails, and it is therefore generally accepted that they belong to the freeze-avoiding species^{5,6}. As the name of this overwintering strategy indicates, these species have physiological adaptations enabling them to avoid internal ice formation even though the ambient temperature for long periods can be much lower than the melting point of their hemolymph. However, springtail species have followed two fundamentally different trajectories of adaptation to subzero survival by freeze-avoidance. Species living predominantly on the ground surface or in vegetation (epigeic species) have relatively impermeable cuticula and are typical freeze-avoiders with high capacity for supercooling similar to many insects^{6,7}. Other springtails inhabiting deeper layers of the soil (hemi- and eu-edaphic species), have little cuticular resistance to desiccation, and base their freeze-avoiding capacity on an alternative strategy termed cryoprotective dehydration. In this strategy, the difference in water vapor pressure between ice in the soil and the supercooled hemolymph drives a net outflux of water vapor^{8,9}. The force of this vapor pressure difference is so large that even a few degrees of supercooling will result in substantial water loss, continuing until the vapor pressure of body fluids equals that of the surrounding ice^{8,10}. At this stage, the risk of ice formation in the body has been eliminated, and subzero survival is ensured.

Studies have shown that *M. arctica* on a seasonal timescale efficiently adjusts the melting point of its hemolymph to equal the temperature of its winter habitat⁸. However, in the initial phases of cryoprotective dehydration, melting point depression rates can be slow with the result that hemolymph at this stage is supercooled by a few degrees until water contents have decreased to levels where further evaporative water loss produces higher rates of melting point depression¹¹. Even though supercooling is limited to a few °C, and of short duration, physical contact between springtails and ice crystals in the habitat potentially may result in inoculative spread of external ice to hemolymph through hydrophilic surfaces or openings of the animal such as mouth or ventral tube.

¹Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada. ²Section of Terrestrial Ecology, Department of Bioscience, Aarhus University, Vejløvej 25, 8600, Silkeborg, Denmark. ³Arctic Research Center, Aarhus University, Ny Munkegade 114, 8000, Aarhus C, Denmark. *email: daviesp@queensu.ca

These considerations prompted us to look for antifreeze proteins (AFPs) which are well-known for their ability to limit ice growth in freeze-avoiding fish and insects^{12–14}.

An AFP was previously characterized from a springtail species, *Hypogastrura harveyi*, present in Eastern Ontario¹⁵. Extracts of active *H. harveyi*, collected from snowbanks in late winter, were able to depress the freezing temperature by 5.8 °C below the melting point. The proteins responsible for this strong thermal hysteresis (TH) activity were unlike any previously described AFPs from fish, insects, plants and microorganisms, being rich in glycine and having tandem tripeptide repeats of Gly-aa2-aa3, where aa2 was frequently also Gly. As predicted by Lin *et al.*¹⁶, these glycine-rich tracts folded as polyproline type II helices linked by turn regions¹⁷. The small (6.5 kDa) isoform of *H. harveyi* AFP (*HhAFP*) had six of these helices forming two sheets secured by intrachain hydrogen bonds in the glycine-rich core, and by two disulfide linkages. A larger isoform (15.7 kDa) was predicted to fold in a similar manner but with 13 polyproline type II helices, again packed into two sheets with a glycine-rich core¹⁸.

Another reason to probe *M. arctica* for AFPs is the expectation of finding additional novel ice-binding proteins in different Collembola families that might elucidate structure–function relationships in AFPs and the mechanism by which these proteins bind to ice¹⁹. Four different AFP types have been found in teleost fishes¹⁴. Several different types have been characterized in insects; and those described to date in plants and microorganisms have added to the remarkable diversity of ice-binding proteins. It is thought that the variety in fish AFPs reflects the relatively recent impact of sea-level glaciation in the Cenozoic era that challenged teleosts to develop ice control at a late stage of their radiation when extant families already existed^{20,21}. Collembola are considerably older than teleost fishes, widespread on all continents, including regions where there are sub-zero temperatures, and might therefore yield a similar diversity of AFPs. Indeed, preliminary characterization of *Gomphiocephalus hodgsoni* from Antarctica suggests it contains an AFP with a distinct amino acid composition that is rich in Cys and His²².

Results

Megaphorura arctica contains a hyperactive AFP. An initial extraction of *M. arctica* (2.3 mg) with buffer (18.4 µL) categorically showed the presence of an AFP in the crude homogenate based on ice crystal shaping and high thermal hysteresis activity (Fig. 1). The ice formed into a distinctive shape with rounded prism surfaces that tapered to points at the basal planes as it was melting (Fig. 1A). This shape was retained unchanged as the temperature was decreased from just under the melting point (Fig. 1B) until 3.1 °C below the melting point (Fig. 1C). When this depressed freezing point was finally exceeded, the ice crystal grew explosively, forming dendritic arms that erupted at angles slightly above and below the plane perpendicular to the *c*-axis linking the crystal tips (Fig. 1D). These characteristics are indicative of hyperactivity and are typical of arthropod AFPs^{23,24}.

The larger extraction of freeze-dried *M. arctica* (0.3 g) with buffer (50 mL) was performed to obtain enough AFP for biochemical characterization. The AFP was isolated through four cycles of rotary ice-affinity purification with the emphasis placed on purity rather than yield. Thus, the ice shells were rinsed before being melted and re-extracted; and no back-extractions were performed on the excluded liquid fractions to recover any AFP that was left behind. When the concentrated final ice fraction (500 µL) was tested, it had a thermal hysteresis activity of 3.9 °C, with ice crystal shaping and growth characteristics (Fig. 1E–H) that were similar to those of the original crude extract (Fig. 1A–D). However, in this latter example the crystal orientation happened to be rotated almost 90°, such that the *c*-axis was perpendicular to the image. The six prism surfaces around the circumference of the crystal were evident during melting (Fig. 1E) and the crystal did not visibly change throughout the thermal-hysteresis gap (Fig. 1F–G). When the freezing point was exceeded, the dendritic burst pattern clearly showed the crystal's hexagonal symmetry with ice growth roughly perpendicular to the *c*-axis (Fig. 1H). When the purified protein was tested at lower concentrations, the crystal still melted into a shape with rounded prism surfaces (Fig. 1I) and the burst was *a*-axial but non-dendritic (Fig. 1J) as the solution was only supercooled by 0.11 °C rather than >3 °C when freezing began. The hexagonal nature of a held crystal and the burst was also evident when the *c*-axis was perpendicular to the page (Fig. 1K,L).

MaAFP comprises several small isoforms with high-glycine content. Analysis of the concentrated pure extract by MALDI mass spectrometry showed the presence of several low molecular weight proteins (Fig. 2). The first large peak included masses of 6518 Da, 6569 Da and 6592 Da, followed by a second large peak of 7112 Da, all with uncertainties of ±5 Da. There was also a smaller peak at 16887 Da ±10 Da. The peaks at 3354 and 8448 *m/z* are likely doubly-charged versions of the 7112-Da and 16887-Da species respectively. The amino acid composition of the extract (Table 1) was highly enriched in Gly (36 mol%), with Ala (13 mol%) being the next most abundant amino acid. This bias in composition was reminiscent of *H. harveyi* AFP (*HhAFP*), where the 6.5-kDa isoform had 48 mol% Gly and 14 mol% Ala. The dominance of Gly and Ala also suggested that the several major peaks seen in the MALDI mass spectrum are isoforms of *MaAFP* rather than impurities. Both purified *MaAFP* and the *HhAFP* isoforms had few of the larger amino acids, such as the aliphatic residues (Ile, Leu, Met) and aromatic residues (Phe, Tyr). The somewhat higher combined level of these residues (18%) and lower level of Gly in the ice affinity-purified *MaAFP* material, compared to the defined sequences, could be due to small amounts of contaminating proteins.

MaAFP resembles HhAFP in sequence and structure. The presence of both lysine and arginine in the *MaAFP* amino acid composition (5 mol% total) encouraged us to digest the protein with trypsin and sequence the resulting peptides (Supplementary Fig. S1) by tandem mass spectrometry (MS/MS). Only one (EW753059) of over 16,000 *M. arctica* expressed sequence tags (ESTs) in the NCBI database²⁵ showed significant similarity to any of the MS/MS sequences (Fig. 3). The open reading frame in frame 3 of this DNA sequence (shown in bold) coded for an 87-residue secreted protein (after removal of the signal polypeptide) with a predicted whole mass of 6516.8 Da, which matched the lightest species in the first large peak of the MALDI spectrum (6518 Da). Moreover,

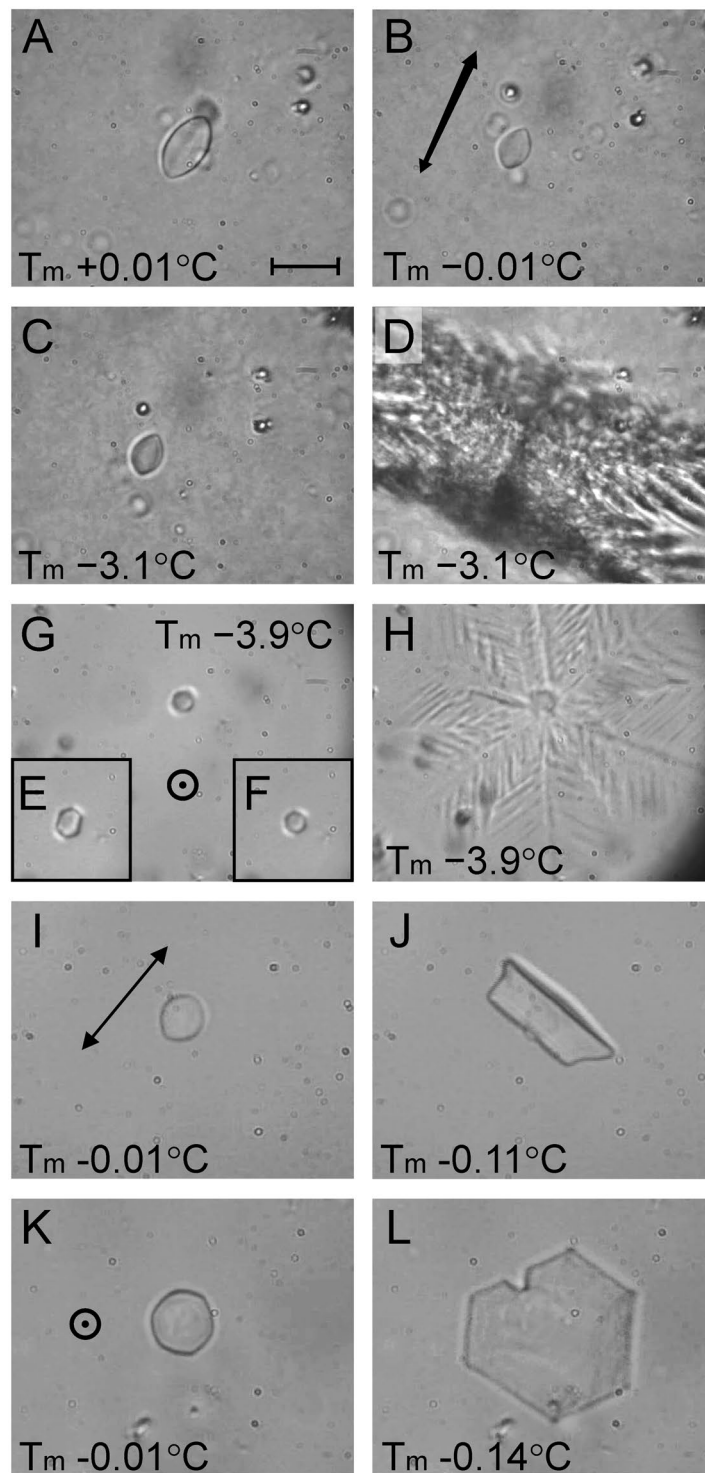


Figure 1. Crystal shaping and thermal hysteresis activity of crude *M. arctica* homogenate (A–D) and purified native *MaAFP* (E–H). (A) Crystal shaping attained during melting of monocrystalline ice at 0.01 °C above the melting point (T_m). The scale bar is 20 μm and is the same for all images. (B) Final shape of crystal at the start of cooling. The double-headed arrow shows the orientation of the *c*-axis. (C) Crystal shape one frame (1/30 s) prior to the burst at 3.1 °C of supercooling. (D) First frame once freezing was initiated. (E–H) Measurements taken on pure *MaAFP* as described in (A–D) except that the *c*-axis is perpendicular to the page and the TH was 3.9 °C. The circle with a central dot indicates the ice *c*-axis is normal to the page. (I–L) Ice crystal shaping attained with diluted purified *MaAFP* prior to its concentration, with the *c*-axis parallel (I–J) or perpendicular (K–L) to the page.

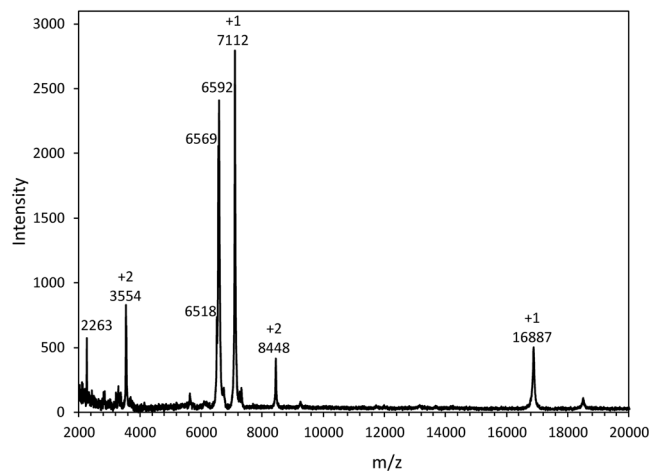


Figure 2. MALDI MS of purified AFPs following four rounds of ice-affinity purification. Peak intensity is plotted against the mass to charge ratio (m/z). Singly (+1) and doubly (+2) charged peaks are indicated.

Amino Acid	Purified <i>Ma</i> AFPs	<i>Ma</i> AFP	<i>Hh</i> 6.5 kDa	<i>Hh</i> 15.7 kDa
Asx	9	6	9	7
Ser	7	6	4	5
Gly	36	49	48	51
Thr	4	1	4	3
Ala	13	22	14	16
Pro	4	8	8	4
Val	5	4	5	2
Lys	3	4	4	2
Arg	2	0	1	1
Other	18	0	3	7

Table 1. Molar percentages of amino acids (totaled to 100% with Cys excluded) in various springtail AFPs. The composition of the purified *Ma*AFPs was determined by amino acid analysis, whereas the compositions of specific isoforms were derived from EST/cDNA sequences (Supplementary Table S1). “Other” includes Met, Leu, Ile, Glx, Tyr, Phe and His. Trp was absent or not detected.

the mass and sequence of the three larger (>500 Da) tryptic peptides predicted from the mature sequence (all except GGAAGK) were exactly matched by tryptic MS/MS sequences. In addition, the majority of the polymorphisms observed in the second and fourth fragments, such as A to P, G to L and K to R, would increase the mass of the protein and may explain the peaks at 6569 to 7112 Da. Partial sequences of several other Gly-rich MS/MS sequences from the ice-affinity purified AFPs (not shown) did not align well with EST EW753059, and may correspond to the 16.9-kDa species.

The sequence of *Ma*AFP contained six short tracts of the tripeptide repeat Gly-aa2-aa3, and in three of these tracts aa2 was also Gly. These sequences and their distribution in the protein are reminiscent of *Hh*AFP. When these two small AFPs were compared they had 62.4% identity, 68.8% similarity and only three gaps in the alignment where *Ma*AFP has additional residues (Fig. 4). The two sequences are aligned to emphasize the Gly-rich tracts linked by more variable turn regions. In both proteins the Gly-rich tracts with the second Gly in the tripeptide repeats were the third, fourth and fifth ones in the sequence. All three gaps in the alignment occur in, or next to, the turns. Also, all four Cys are conserved in their locations.

Based on the conserved sequence features mentioned above, it was relatively simple to homology model the three-dimensional fold of *Ma*AFP (Fig. 5). The Phyre 2 web-based system²⁶ for protein structure prediction folded the six Gly-rich tracts as polyproline type II helices and packed them into a bundle based on the *Hh*AFP structure. The Phyre2 model had >99% confidence over 93% of the sequence of *Hh*AFP. This model was based on the *Hh*AFP structures solved by Pentelute *et al.*¹⁷ In this initial model (not shown), helix 4 was rotated relative to the final model with the Ala sidechains directed inwards and the Gly residues surface exposed. This was easily corrected using the program Modeller so that all residues larger than Gly, within the six helices, are now surface exposed (Fig. 5A). The only exceptions to this are the Ser found near the end of helix 4 and the Ala in the second loop of helix 5 (Fig. 4, pink highlighting). The Ser sidechain is easily accommodated as it is adjacent to the C-terminal residue of helix 6. The sidechain of the Ala in helix 5 (which is a Gly in *Hh*AFP) is partially exposed and a small (0.4 Å) shift of the backbone was adequate to prevent steric clashes (Fig. 5A,B). The stability of this

```

gcSGAPGAPGANGADGCPGGVGSAGSVG--GPGCDGGAGGNGGK 3442.39  A
GCSGAPGAPGANGADGCPGGVGSAGSVG--GPGCDGGAGGNGGK 3442.42
                                     ggDGGAGGV 2612.16 #1
      ???GADGCPGSVGSAGS AADAGSGCDGGNGGNGGK 3434.32 #4
      ???GAAGCPGSVGGAGS AAAAGSGCHGGR 3011.25 #3

      gtGPACAGGAGGAGGASGGGGGK 1688.73  B
      GTGPACAGGAGGAGGASGGGGGK 1688.74
      DPAGAGPCAGGAGGAGGASGGLGSR 2612.16 #1 (continued)
      PAGAGPCAGGAGGAGGASGGLGSR 1869.89 #7
      gtGPACAGGAGGAGGASGGAAGGK 1702.75 #2

GGAAGK 460.24 n.d.

GGAGSPPGADGAVGAA 1211.57  C
GGAGSPPGADGAVGAA 1211.57
ggAGSPPGADGAPGAP 1234.56 #6
ggAGSPPGADGPPGAP 1261.57 #5

```

Figure 3. Translation of *MaEST* EW753059.1 aligned with sequences determined by tandem mass spectrometry sequencing of tryptic fragments from purified *MaAFP*. The 87-residue open reading frame from the EST (with the signal sequence removed) is shown sequentially in bold as four tryptic fragments of masses 3442.42, 1688.74, 460.24 and 1211.57. Gly residues predicted to form the core are highlighted orange. Above it are the tandem MS/MS sequences that exactly correspond to and support the EST sequence. Below it are homologous fragments producing some of the higher total ion counts (excluding trypsin autolysis fragments) numbered #1 to #7 shown in Supplementary Fig. S1. Differences in these sequenced fragments relative to the bolded EST translation are highlighted yellow. Residues that were inferred from b2 ion masses are in lowercase, and undetermined regions of unknown length are represented by question marks. The predicted (EST) and observed (MS/MS fragment) monoisotopic H^{+1} masses (in Da) are indicated on the right.

model was assessed by molecular dynamics over a 20-ns timespan (Supplementary Fig. S2). The RMSD remained relatively stable with an average deviation of 2.1 Å from the starting point, which was only 0.8 Å higher than that found when the known *HhAFP* structure was used in the simulation. In both instances, the polyproline helices retained their relative positions and the core Gly residues remained buried.

A comparison of the *MaAFP* model with the *HhAFP* structure shows that the two proteins have the same amphipathic nature, where the outer surface formed from the three even-numbered 2nd, 4th and 6th helices (Fig. 5A) is relatively hydrophobic because of its high alanine content (Fig. 5B, white residues). This ice-binding site is almost identical between the two proteins and is dominated by out-facing Ala (white), Val (light grey) and Pro (dark grey) residues, where all of the Gly (orange) face inward. The only difference is that one of the Val residues (light grey) is in a different position. The opposite outer surface formed by the odd-numbered helices is more hydrophilic and contains several charged residues like Lys and Asp (Fig. 5C). This is the non-ice-binding site exposed to the bulk solvent when the AFP is bound to the ice surface, and here the residues are not nearly as well conserved. This surface is slightly less hydrophilic in *MaAFP* as there are several substitutions of charged or polar residues with Ala (white) or Pro (dark grey). The turns at both ends are also somewhat variable, but the two disulfide bonds (Fig. 5B,C) are conserved: linking the start of helix 1 to the end of helix 2, and end of helix 1 to the end of helix 3.

The untranslated regions (UTRs) of *HhAFP* and *MaAFP* are highly divergent. To probe the evolutionary relationship between *HhAFP* and *MaAFP* we performed dot matrix analyses comparing the cDNAs encoding the AFPs (Fig. 6). For controls we did similar comparisons on two orthologous proteins from these two springtail species. Each 20 bp segment of sequence 1 (sliding in 1 bp increments) was compared to every 20 bp segment of sequence 2. Matches above the threshold are indicated by diagonal lines, which show the relative positions of matches in both sequences.

When cDNAs encoding two well-conserved orthologous proteins (eukaryotic translation initiation factor 4E (EIF4) and ribosomal protein L35 (RPL35)) were compared, the diagonal lines spanning the coding regions indicate that these sequence segments are indeed well conserved (Fig. 6A,B). However, the lack of continuous diagonal lines across the 5' and 3' UTRs suggest that the non-coding segments are not conserved. When a similar analysis was done with the AFP sequences of the two species, multiple matches, which appear as numerous parallel diagonal lines, were found between the coding regions of the mature AFPs (Fig. 6C). These diagonals reflect the similarities between the Gly-rich repeats, but the plot lacks the strong continuous diagonal indicative of well-conserved orthologues. In addition, the UTRs and the region encoding the signal peptide are not conserved. A plot comparing the intraspecific large and small AFP-encoding cDNAs of *H. harveyi* showed the same pattern (Fig. 6D). One interesting similarity between the UTRs of all these genes was that they had a reduced GC content

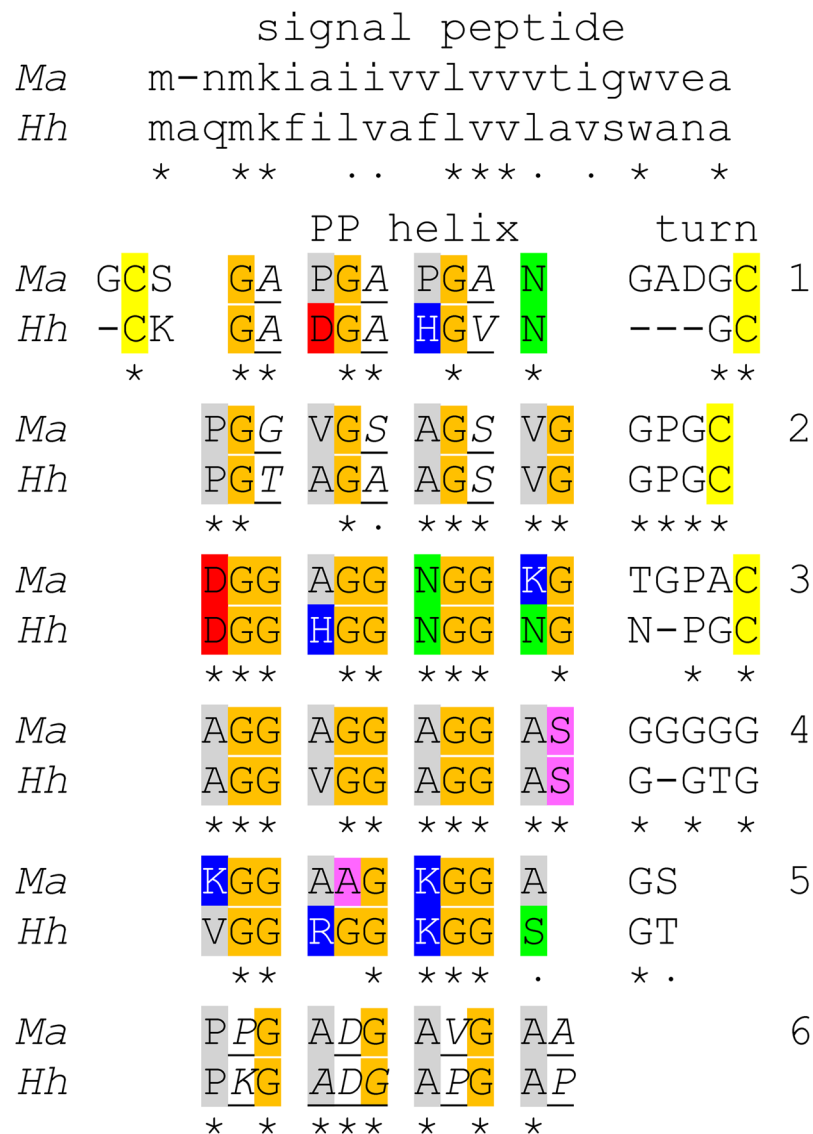


Figure 4. Alignment of *Ma*AFP and *Hh*AFP. Gly residues within the polyproline type II helices that form the core of the proteins are shaded orange. Residues exposed on the ice-binding surface (helices 2, 4, 6) and backside (helices 1, 3, 5) are highlighted grey (hydrophobic), red (acidic), blue (basic) or green (polar). Cys are highlighted yellow; residues with sidechains that are exposed on the sides of the protein are in italics and underlined, and the non-Gly residues within the core are highlighted pink. Identity and similarity are denoted by asterisks and dots, respectively. Signal polypeptides sequences are in lower case at the top.

(23 to 40%) relative to the coding sequences (46 to 61%). The scattered matches between the 3' UTRs of *Ma*- and *Hh*-EIF4 (Fig. 6A) correspond primarily to A-rich tracts within these AT-rich sequences and are not indicative of homology.

Discussion

Examples of highly potent AFPs are mostly found in freeze-avoiding insects with high supercooling capacity^{27–30}. For instance, AFP-mediated freezing point depression in the spruce budworm (*Choristoneura fumiferana*) of 6°C is an important component in overwintering of first-instar larvae remaining unfrozen and viable in the boreal forest at temperatures of –20 to –30°C²⁸. Despite *M. arctica*'s reliance on cryoprotective dehydration, and not deep supercooling, as its main strategy for avoiding freezing, these springtails also possess a potent AFP that complements the set of adaptive traits for winter survival in this species.

Due to logistic limitations we were not able to measure antifreeze hysteresis activity of hemolymph samples from live specimens in the field, and we therefore do not have a direct determination of the amount of freezing point depression *Ma*AFP can afford its host. However, since we know the dry weight of the animal sample and how much buffer was used for the crude extract measurements, we can make an estimate of this. We made an extract from 2.3 mg dry *M. arctica* tissue in 18.4 µL of buffer, equivalent to a water content of ca. 7.8 mg water mg⁻¹ dry tissue, which is at least twice the normal water content of fully hydrated *M. arctica*^{31,32}. It is therefore

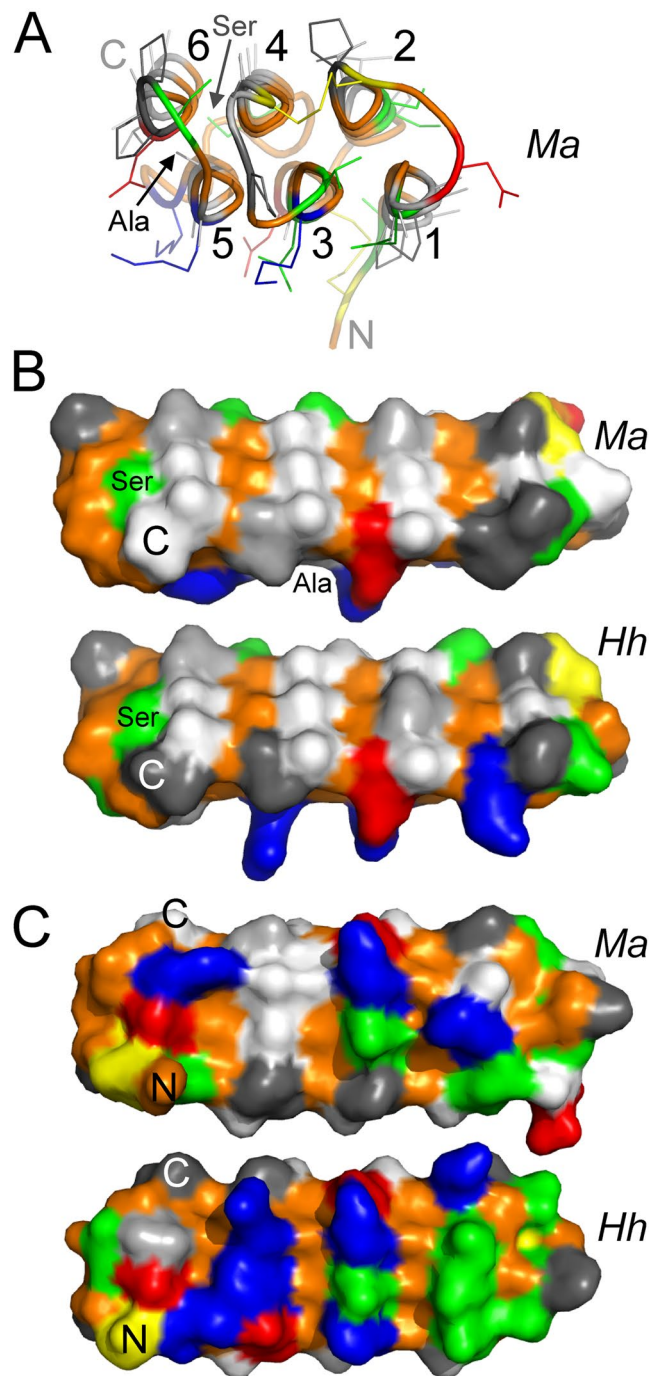


Figure 5. Comparison of the *Ma*AFP model and *Hh*AFP structure. (A) End-on view of *Ma*AFP displayed in PyMOL⁶⁸ with the backbone rendered in cartoon mode and sidechains (no H) as sticks, colored as below except Ala were rendered in very light grey. (B) Oblique surface rendering of the ice-binding surfaces of *Ma*AFP (top) and *Hh*AFP (bottom). Residues are colored as follows; Ala-white, Val-light grey, Pro-dark grey, Cys-yellow, Gly-orange, basic-blue, acidic-red, polar-green. N- and C-termini are indicated. (C) Face-on views of the back surface opposite to the ice-binding surface.

likely that the freezing point depression provided by *Ma*AFP *in vivo* is similar or perhaps even higher than the 3.1 °C measured for the crude extract.

Since cryoprotective dehydration can sometimes be a slow process some supercooling capacity (typically 2–4 °C) is needed in the initial phases of dehydration, especially if the cooling rate is high^{8,31,33}. At this stage, there is a risk that external ice will spread to the animal causing inoculative freezing. Until now, we have been missing a good explanation how inoculative freezing is avoided in frozen soil, but the discovery of a potent AFP in *M. arctica* closes this gap in knowledge. Thus, the freezing point depression that we estimated to ca. 3 °C seems to meet with the needs outlined above. Studies of the exact location of *Ma*AFP in the animal would further explain

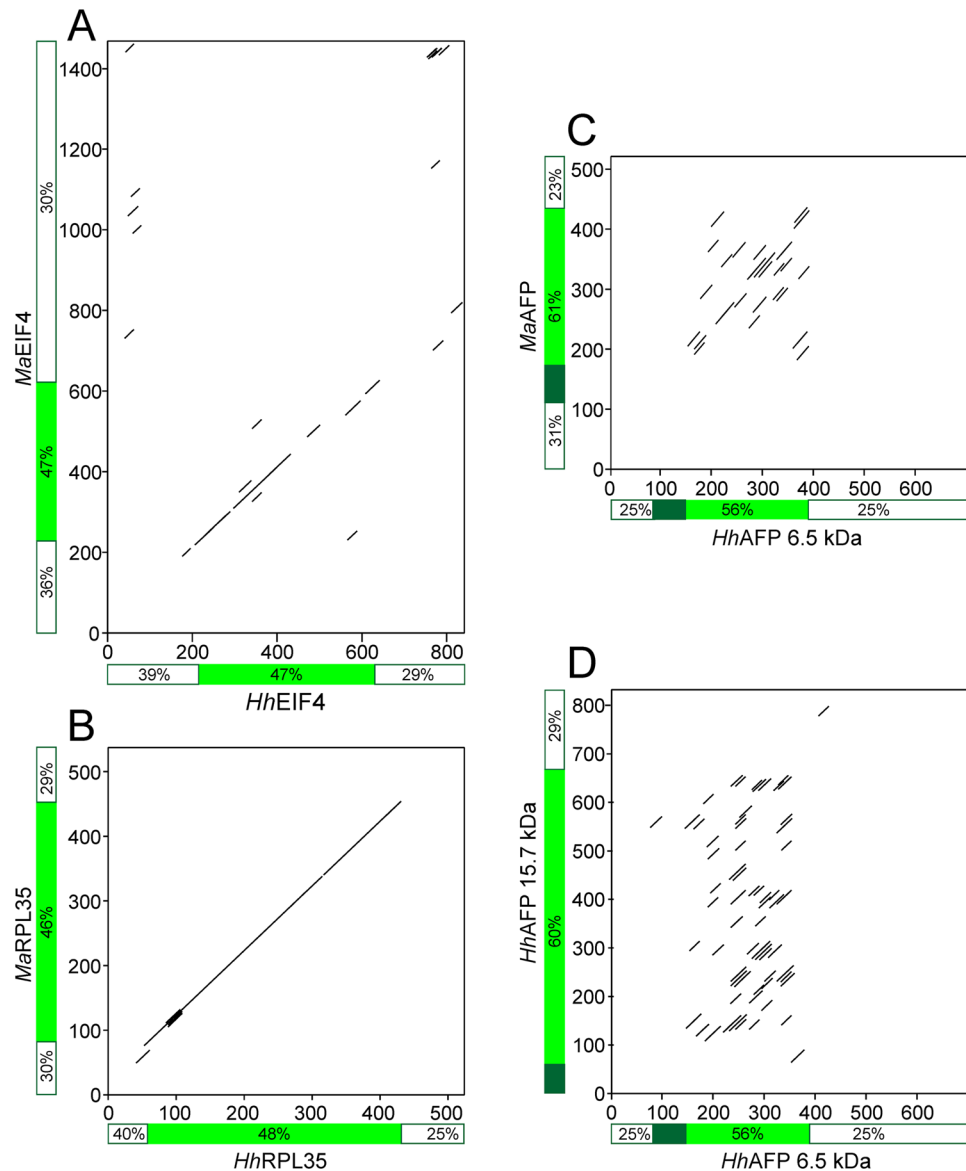


Figure 6. Dot plot comparison of cDNA sequences encoding orthologous proteins (**A,B**) or AFP sequences (**C,D**) from *M. arctica* and *H. harveyi*. A diagonal line indicates that the two sequences match with a threshold value over 50%, scored across a 20-bp sliding window. All axes are in bp with the 5' end of each cDNA beginning at zero. The green bars indicate the position of coding sequence with the signal peptide in dark green (if present). The percent GC in each region (5' UTR, coding region, 3' UTR) is indicated alongside each axis. Note that the 5' UTR of *HhAFP* 15.7 kDa was not sequenced and the green bars do not overlay any diagonal lines. Accession numbers are given in Supplementary Table S1.

the ability to avoid inoculative freezing. AFP can be expressed in various tissues of insects such as the midgut, hemolymph and epidermis^{34,35}. Expression of AFP in epidermis cells could be particularly relevant for inoculative freezing in *M. arctica* and other springtails using cryoprotective dehydration³⁴.

Inoculative freezing may be a significant cause of lethality for springtails and many other soil living invertebrates that do not tolerate freezing of their body fluids. For example, Sømme and Conradi-Larsen (1977) observed that cooling the springtail *Tetracanthella wahlgreni* to -3°C in close contact with ice caused inoculative freezing and death even though supercooling points of this species ranged between -15 and -25°C if specimens were not in contact with external ice³⁶. Since contact with ice is decisive, the risk of inoculative freezing is highest when soil moisture content is high^{37,38}. Interestingly, several hygrophilic soil invertebrates (e.g. nematodes, enchytraeids and mites) survive sub-zero temperatures by freeze-tolerance, or can switch between cryoprotective dehydration and freeze-tolerance depending on the moisture conditions of their overwintering habitat^{39–41}. This dual strategy is consistent with the fact that both freeze-tolerance and cryoprotective dehydration involves tolerance of cellular dehydration and the physiological adaptations connected to this¹¹.

MaAFP has the attributes of a hyperactive antifreeze protein in having high thermal hysteresis activity at low protein concentrations, with evidence of AFP binding to the basal plane. This is typical of arthropod AFPs, which have generally been selected for higher freezing point depression activity than the moderately active AFPs in fish, which never face temperatures lower than -1.9°C ^{23,24}. *MaAFP* is made up of what appear to be several isoforms. Multiple AFP isoforms are common in fish^{42–45} and arthropods^{35,46,47}. They arise from gene duplication followed by divergence. Since freezing point depression is a function of AFP concentration, AFP-producing organisms seem to be under selective pressure to increase the amount of circulating AFP by gene amplification^{42,48}. For AFPs made from repetitive sequences it is easy to generate length variation by expansion and/or contraction of the repeats. Examples of this can be seen in AFPs from the beetles *Tenebrio molitor* and *Dendroides canadensis* where the repeating unit is a 12-residue coil in a beta-solenoid structure. Common isoforms have seven coils but some have eight^{46,49}. Another example of length variation is in the moth, *Choristoneura fumiferana*, where one isoform has two extra 15-residue coils in its beta-solenoid structure⁵⁰. Interestingly, these larger isoforms are more active than the shorter ones because the ice-binding site is increased in area. Thus, there may also be selective pressure to increase isoform size.

Lengthening of the polyproline type II coils in *HhAFP* and *MaAFP* seems feasible too, since they are based on Gly-aa2-aa3 tandem tripeptide repeats. However, unlike the single solenoid structures of the beetle and moth AFPs, the polyproline type II coils must pack into a higher order tertiary structure where the coils are bundled together to form a Gly-rich core. Thus, the expansion of a single coil would potentially destabilize the overall structure. Instead, an increase in isoform size in *HhAFP* seems to have been accommodated by an increase in the number of coils from 6 to 13, and with it again a considerable increase in antifreeze activity¹⁸.

Despite the high sequence identity between *MaAFP* and *HhAFP* it is not possible to unequivocally say they are descended from a common ancestor sequence. There are well-documented examples of simple repetitive AFPs that have independently originated from different progenitors. For example, the antifreeze glycoproteins (AFGPs) of Antarctic Notothenioids and Arctic cods are made up of a tandemly repeated Ala-Ala-Thr tripeptide with a disaccharide O-linked to the Thr^{44,45}. Although these AFGPs are virtually indistinguishable and share almost complete sequence identity as a set of isoforms of different lengths derived from polyprotein precursors, they are not homologs. The Notothenioid AFGPs have arisen from an intronic sequence from the gene for the pancreatic enzyme, trypsinogen^{51,52}. AFGPs of the northern cods, however, are derived from a different DNA sequence⁵³. Type I AFPs are another example of convergent evolution to a structural and functional commonality. These Ala-rich, alpha-helical AFPs, with an underlying 11-amino-acid repeat, have independently arisen at least four times in different branches of teleost fishes from unknown precursor sequences⁵⁴. Since they are 60–65% Ala they all have >25% sequence identity, which is the measure applied to recognize homology. Clearly this criterion is invalid when dealing with proteins that have such a biased amino acid composition.

The comparison between *MaAFP* and *HhAFP* is more complex because, unlike the AFGP and type AFP I examples above, these AFPs have a higher order of protein structure where six helices of similar length pack together to form a Gly-rich core. In addition, there are four Cys, whose positions are conserved such that the two disulfide bonds they form to help secure the helix bundle structure are in the same places in both proteins. These structural features argue for homology. A comparison of the 5'- and 3'-untranslated regions shows no sign of sequence identity, which could argue against homology. One possible explanation for these confounding results is that *MaAFP* and *HhAFP* are ancient homologs that have survived from an earlier ice age, perhaps even before these two species diverged. Collembola originated at least 400 Ma and would have encountered the ice ages of the Permian period ~300 Ma^{55–58}. If so, some of the extant springtails may have derived their AFPs from this period. Those that retained their AFP genes during the 200+ million years between the Permian and Cenozoic ice ages would have been well equipped to survive in present sub-zero conditions. This hypothesis might explain why the UTRs of the *HhAFP* large and small isoforms also appear unrelated, as do the UTRs of *MaAFP* and *HhAFP*. If the isoforms predate speciation and date back to the Permian period, regions of the gene not under selective pressure would have had ~300 million years to drift. One way to test this hypothesis would be to sample additional freeze-resistant springtail species for the type of AFPs they might possess to see if they too have high-Gly, polyproline type II helical bundles suggestive of an early origin.

Materials and Methods

Collection and storage of springtails. Adult *M. arctica* were collected in August 2017, on a beach near Sauðárkrúkur, north-west Iceland, and brought back to the laboratory in Silkeborg, Denmark. The animals were kept at 5 °C in 200-mL beakers on a substrate of decomposing brown algae. In order to acclimate springtails to winter conditions, they were kept in darkness at 5 °C for three months followed by about two months at 1.5 °C. At this point, ~1,500 adult animals were taken and freeze dried for two days to complete dryness. The dried animals were stored in Eppendorf vials at -80°C until they were shipped by courier to Queen's University, Canada.

Antifreeze activity measurements. Antifreeze activity, referred to here as thermal hysteresis (TH), was measured in a nanolitre osmometer⁵⁹ as the depression of the freezing point (°C) below the melting temperature. To establish the presence or absence of TH, a soluble extract was prepared from freeze-dried *M. arctica* (2.3 mg). The dried tissue was mixed with 18.4 µL of homogenization buffer (50 mM Tris-HCl (pH 7.8), 100 mM NaCl, 2 mM phenylthiocarbamide, 1x cOmplete, EDTA-free protease inhibitor tablet (Roche, Mississauga, Canada)) and homogenized by hand, using a pellet pestle (Sigma, Burlington, MA) in a 1.5-mL tube. The homogenate was centrifuged at $16,000 \times g$ for 15 min and the supernatant tested for antifreeze activity.

Ice-affinity purification of AFP. For a larger-scale extraction, freeze-dried *M. arctica* (0.30 g) were homogenized at 4 °C in 50 mL of the above buffer containing 1 mM EDTA. The homogenate was centrifuged at $31,000 \times g$ for 30 min, and the supernatant was filtered through glass wool to remove lipid. AFP was extracted from the

filtered supernatant (<50 mL) using four cycles of rotary ice affinity purification^{60,61}, but scaled down in size from the previously reported protocol. Thus, the starting ice shells contained 10–15 mL of ddH₂O in a 250-mL round bottom flask. The rotating ice-shells were cooled in an ethylene glycol bath for 1.5–2 h starting at –1.5 °C and decreased by 0.2 °C every 0.5 h until ~25 mL of liquid remained in the flask. Once the liquid was ~50% incorporated into the ice-shell the 250 mL flask was removed from the apparatus, the liquid was poured out and diluted to 50 mL with ddH₂O, and the ice-shell was quickly rinsed with 5 mL of homogenizing buffer. The ice shell was melted with 5 mL of 10× homogenate buffer containing 10 mM PTU and 10 mM PMSF then diluted to 50 mL with ddH₂O. This process was repeated for four rounds of ice-affinity purification. The final melted ice sample was concentrated using an Amicon Ultracel 3 K Centrifugation filter spun in a refrigerated Sorvall ST16R centrifuge at 4000 rpm until a volume of ~500 µL was reached.

Amino acid analysis. Amino acid composition analysis was performed at the SickKids Proteomics, Analytics, Robotics & Chemical Biology Centre (SPARC) BioCentre as previously described⁶¹.

Mass spectrometry. MALDI mass spectrometry was performed using α-cyano-4-hydroxycinnamic acid matrix on dried droplets with a SCIEX Voyager DE Pro in Linear mode by the Queen's Protein Function Discovery Facility. Tandem mass spectrometry protein sequencing of tryptic fragments separated by high-performance liquid chromatography (LC-MS/MS) was done using ESI (nanospray) with an Orbitrap (FT-ICR) by SPARC BioCentre Molecular Analysis (The Hospital for Sick Children, Toronto, Canada).

Transcriptome data analysis. *H. harveyi* AFP sequences (ABB03725.1 and ADI39904.1) were used as query sequences in BLAST searches (NCBI, with compositional adjustments and low-complexity filters off) of *M. arctica* sequences⁶². Additionally, all 16379 *M. arctica* ESTs were downloaded from NCBI, and translations of all six reading frames⁶³ were scanned for sequence strings identified by LC-MS/MS. To find control sequences for dot blot comparisons, previously sequenced *H. harveyi* ESTs for nuclear genes were searched by BLAST⁶⁴ for two examples that encoded homologous sequences in the *M. arctica* EST database. These were then assembled into contigs using the Cap3 Sequence Assembly Program⁶⁵. Dotmatcher <http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher>, written by Ian Longden, formerly at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK, was used to compare dot plots of the two nuclear gene cDNAs with complete coding sequences, 5' UTRs ≥ 60 bp and complete 3' UTRs, as well as cDNAs encoding the AFPs.

Molecular modeling. The *Ma*AFP sequence was submitted to Phyre2²⁶ using both normal and intensive mode. An alignment sequence of *Ma*AFP against *Hh*AFP was inputted into MODELLER which performed homology modeling to produce the refined model of *Ma*AFP⁶⁶. GROMACS was used with both the refined model of *Ma*AFP and the crystallographic structure of *Hh*AFP⁶⁷. The structures were energy minimized after being solvated in a box of water, after which 0.1 ns position-restrained molecular dynamics runs at constant-volume and constant-pressure were done. These were followed by unrestrained molecular dynamics at a temperature of 277 K for 20 ns where temperature and pressure were maintained with the v-rescale and Parrinello–Rahman protocols, respectively.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 24 December 2019; Accepted: 31 January 2020;

Published online: 20 February 2020

References

- Fjellberg, A. *The Collembola of Fennoscandia and Denmark. Part I: Poduromorpha*. (Brill, Leiden, 1998).
- Hodkinson, I. D. *et al.* Feeding studies on *Onychiurus arcticus* (Tullberg) (Collembola: Onychiuridae) on West Spitzbergen. *Polar Biology* **14**, 17–19 (1994).
- Bahrndorff, S., Petersen, S. O., Loeschke, V., Overgaard, J. & Holmstrup, M. Differences in cold and drought tolerance of high arctic and sub-arctic populations of *Megaphorura arctica* Tullberg 1876 (Onychiuridae: Collembola). *Cryobiology* **55**, 315–323, <https://doi.org/10.1016/j.cryobiol.2007.09.001> (2007).
- Coulson, S. *et al.* Thermal environments of arctic soil organisms during winter. *Arctic and Alpine research* **27**, 364–370 (1995).
- Holmstrup, M. Screening of cold tolerance in fifteen springtail species. *J. Therm. Biol.* **77**, 1–6, <https://doi.org/10.1016/j.jtherbio.2018.07.017> (2018).
- Cannon, R. J. C. & Block, W. Cold tolerance of microarthropods. *Biological Reviews* **63**, 23–77 (1988).
- Somme, L. Supercooling and winter survival in terrestrial arthropods. *Comp. Biochem. Physiol.* **73A**, 519–543 (1982).
- Holmstrup, M., Bayley, M. & Ramlov, H. Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates. *Proc. Natl. Acad. Sci. USA* **99**, 5716–5720, <https://doi.org/10.1073/pnas.082580699> (2002).
- Salt, W. Delayed inoculative freezing of insects. *The Canadian Entomologist* **95**, 1190–1202 (1963).
- Holmstrup, M. & Westh, P. Dehydration of earthworm cocoons exposed to cold: A novel cold hardiness mechanism. *Journal of Comparative Physiology B* **164**, 312–315 (1994).
- Holmstrup, M. The ins and outs of water dynamics in cold tolerant soil invertebrates. *J. Therm. Biol.* **45**, 117–123, <https://doi.org/10.1016/j.jtherbio.2014.09.001> (2014).
- DeVries, A. L. Antifreeze peptides and glycopeptides in cold-water fishes. *Annu. Rev. Physiol.* **45**, 245–260 (1983).
- Duman, J. G. Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annu. Rev. Physiol.* **63**, 327–357 (2001).
- Bar Dolev, M., Braslavsky, I. & Davies, P. L. Ice-Binding Proteins and Their Function. *Annu. Rev. Biochem.* **85**, 515–542, <https://doi.org/10.1146/annurev-biochem-060815-014546> (2016).
- Graham, L. A. & Davies, P. L. Glycine-rich antifreeze proteins from snow fleas. *Science* **310**, 461 (2005).
- Lin, F. H., Graham, L. A., Campbell, R. L. & Davies, P. L. Structural modeling of snow flea antifreeze protein. *Biophys J* **92**, 1717–1723 (2007).

17. Pentelute, B. L. *et al.* X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc.* **130**, 9695–9701, <https://doi.org/10.1021/ja8013538> (2008).
18. Mok, Y. F. *et al.* Structural basis for the superior activity of the large isoform of snow flea antifreeze protein. *Biochemistry* **49**, 2593–2603, <https://doi.org/10.1021/bi901929n> (2010).
19. Davies, P. L. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem Sci* **39**, 548–555, <https://doi.org/10.1016/j.tibs.2014.09.005> (2014).
20. Scott, G. K., Fletcher, G. L. & Davies, P. L. Fish antifreeze proteins: recent gene evolution. *Can. J. Fish. Aquat. Sci.* **43**, 1028–1034 (1986).
21. Cheng, C. H. Evolution of the diverse antifreeze proteins. *Curr. Opin. Genet. Dev.* **8**, 715–720 (1998).
22. Haves, T. C., Marshall, C. J. & Wharton, D. A. A 9 kDa antifreeze protein from the Antarctic springtail, *Gomphiocephalus hodgsoni*. *Cryobiology* **69**, 181–183, <https://doi.org/10.1016/j.cryobiol.2014.07.001> (2014).
23. Scotter, A. J. *et al.* The basis for hyperactivity of antifreeze proteins. *Cryobiology* **53**, 229–239, <https://doi.org/10.1016/j.cryobiol.2006.06.006> (2006).
24. Bar-Dolev, M., Celik, Y., Wettlaufer, J. S., Davies, P. L. & Braslavsky, I. New insights into ice growth and melting modifications by antifreeze proteins. *J. R. Soc. Interface* **9**, 3249–3259, <https://doi.org/10.1098/rsif.2012.0388> (2012).
25. Clark, M. S. *et al.* Surviving extreme polar winters by desiccation: clues from Arctic springtail (*Onychiurus arcticus*) EST libraries. *BMC Genomics* **8**, 475, <https://doi.org/10.1186/1471-2164-8-475> (2007).
26. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845–858, <https://doi.org/10.1038/nprot.2015.053> (2015).
27. Kristiansen, E., Pedersen, S., Ramlov, H. & Zachariassen, K. E. Antifreeze activity in the cerambycid beetle *Rhagium inquisitor*. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology* **169**, 55–60, <https://doi.org/10.1007/s003600050193> (1999).
28. Tyshenko, M. G., Doucet, D., Davies, P. L. & Walker, V. K. The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nat. Biotechnol.* **15**, 887–890 (1997).
29. Bennett, V. A. *et al.* Comparative overwintering physiology of Alaska and Indiana populations of the beetle *Cucujus clavipes* (Fabricius): roles of antifreeze proteins, polyols, dehydration and diapause. *J. Exp. Biol.* **208**, 4467–4477 (2005).
30. Friis, D. S., Kristiansen, E., von Solms, N. & Ramlov, H. Antifreeze activity enhancement by site directed mutagenesis on an antifreeze protein from the beetle *Rhagium mordax*. *FEBS Lett* **588**, 1767–1772, <https://doi.org/10.1016/j.febslet.2014.03.032> (2014).
31. Holmstrup, M. & Somme, L. Dehydration and cold hardiness in the Arctic Collembolan *Onychiurus arcticus* Tullberg 1876. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology* **168**, 197–203, <https://doi.org/10.1007/s003600050137> (1998).
32. Worland, M. R., Grubor-Lajsic, G. & Montiel, P. O. Partial desiccation induced by sub-zero temperatures as a component of the survival strategy of the Arctic collembolan *Onychiurus arcticus* (Tullberg). *Journal of Insect Physiology* **44**, 211–219, [https://doi.org/10.1016/S0022-1910\(97\)00166-2](https://doi.org/10.1016/S0022-1910(97)00166-2) (1998).
33. Elnitsky, M. A., Hayward, S. A. L., Rinehart, J. P., Denlinger, D. L. & Lee, R. E. Cryoprotective dehydration and the resistance to inoculative freezing in the Antarctic midge, *Belgica antarctica*. *Journal of Experimental Biology* **211**, 524–530, <https://doi.org/10.1242/jeb.011874> (2008).
34. Olsen, T., Sass, S., Li, N. & Duman, J. Factors contributing to seasonal increases in inoculative freezing resistance in overwintering fire-colored beetle larvae *Dendroides canadensis*. *J. Exp. Biol.* **201**(Pt 10), 1585–1594 (1998).
35. Duman, J. G., Verleye, D. & Li, N. Site-specific forms of antifreeze protein in the beetle *Dendroides canadensis*. *J. Comp. Physiol. B.* **172**, 547–552, <https://doi.org/10.1007/s00360-002-0284-x> (2002).
36. Somme, L. & Conradi-Larsen, E.-M. Cold-hardiness of collembolans and oribatid mites from windswept mountain ridges. *Oikos* **29**, 118–126 (1977).
37. Costanzo, J. P., Moore, J. B., Lee, R. E. Jr., Kaufman, P. E. & Wyman, J. A. Influence of soil hydric parameters on the winter cold hardiness of a burrowing beetle, *Leptinotarsa decemlineata* (Say). *J. Comp. Physiol. B.* **167**, 169–176, <https://doi.org/10.1007/s003600050061> (1997).
38. Hoshikawa, K., Tsutsui, H., Honma, K. & Sakagami, S. F. Cold Resistance in Four Species of Beetles Overwintering in the Soil, with Notes on the Overwintering Strategies of Some Soil Insects. *Applied Entomology and Zoology* **23**, 273–281, <https://doi.org/10.1303/aez.23.273> (1988).
39. Wharton, D. A., Goodall, G. & Marshall, C. J. Freezing survival and cryoprotective dehydration as cold tolerance mechanisms in the Antarctic nematode *Panagrolaimus davidi*. *J. Exp. Biol.* **206**, 215–221, <https://doi.org/10.1242/jeb.00083> (2003).
40. Pedersen, P. G. & Holmstrup, M. Freeze or dehydrate: only two options for the survival of subzero temperatures in the arctic enchytraeid *Fridericia ratzeli*. *J. Comp. Physiol. B.* **173**, 601–609, <https://doi.org/10.1007/s00360-003-0370-8> (2003).
41. Kawarasaki, Y., Teets, N. M., Denlinger, D. L. & Lee, R. E. Jr. Alternative overwintering strategies in an Antarctic midge: freezing vs. cryoprotective dehydration. *Functional Ecology* **28**, 933–943, <https://doi.org/10.1007/s00300-014-1475-0> (2014).
42. Hew, C. L. *et al.* Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *J. Biol. Chem.* **263**, 12049–12055 (1988).
43. Nishimiya, Y., Sato, R., Takamichi, M., Miura, A. & Tsuda, S. Co-operative effect of the isoforms of type III antifreeze protein expressed in Notched-fin eelpout, *Zoarces elongatus* Kner. *FEBS J.* **272**, 482–492 (2005).
44. Osuga, D. T. & Feeney, R. E. Antifreeze glycoproteins from Arctic fish. *J. Biol. Chem.* **253**, 5338–5343 (1978).
45. Schrag, J. D., O'Grady, S. M. & DeVries, A. L. Relationship of amino acid composition and molecular weight of antifreeze glycopeptides to non-colligative freezing point depression. *Biochim. Biophys. Acta* **717**, 322–326 (1982).
46. Liou, Y. C., Thibault, P., Walker, V. K., Davies, P. L. & Graham, L. A. A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle *Tenebrio molitor*. *Biochemistry* **38**, 11415–11424 (1999).
47. Doucet, D. *et al.* Structure-function relationships in spruce budworm antifreeze protein revealed by isoform diversity. *Eur. J. Biochem.* **267**, 6082–6088 (2000).
48. Scott, G. K., Hew, C. L. & Davies, P. L. Antifreeze protein genes are tandemly linked and clustered in the genome of the winter flounder. *Proc. Natl. Acad. Sci. USA* **82**, 2613–2617 (1985).
49. Andorfer, C. A. & Duman, J. G. Isolation and characterization of cDNA clones encoding antifreeze proteins of the Pyrochroid beetle *Dendroides canadensis*. *J. Insect Physiol.* **46**, 365–372 (2000).
50. Leinala, E. K., Davies, P. L. & Jia, Z. Crystal structure of beta-helical antifreeze protein points to a general ice binding model. *Structure (Camb)* **10**, 619–627 (2002).
51. Chen, L., DeVries, A. L. & Cheng, C. H. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic nototheniid fish. *Proc. Natl. Acad. Sci. USA* **94**, 3811–3816 (1997).
52. Chen, L., DeVries, A. L. & Cheng, C. H. Convergent evolution of antifreeze glycoproteins in Antarctic nototheniid fish and Arctic cod. *Proc. Natl. Acad. Sci. USA* **94**, 3817–3822 (1997).
53. Zhuang, X., Murphy, K. R., Ghigliotti, L., Pisano, E. & Cheng, C. C. Reconstruction of the repetitive antifreeze glycoprotein genomic loci in the cold-water gadids *Boreogadus saida* and *Microgadus tomcod*. *Mar Genomics* **39**, 73–84, <https://doi.org/10.1016/j.margen.2018.02.003> (2018).
54. Graham, L. A., Hobbs, R. S., Fletcher, G. L. & Davies, P. L. Helical antifreeze proteins have independently evolved in fishes on four occasions. *PLoS ONE* **8**, e81285, <https://doi.org/10.1371/journal.pone.0081285> (2013).

55. Dunlop, J. A. & Garwood, R. J. Terrestrial invertebrates in the Rhynie chert ecosystem. *Philos Trans R Soc Lond B Biol Sci* **373**, <https://doi.org/10.1098/rstb.2016.0493> (2018).
56. Baresel, B. *et al.* Timing of global regression and microbial bloom linked with the Permian-Triassic boundary mass extinction: implications for driving mechanisms. *Sci. Rep.* **7**, 43630, <https://doi.org/10.1038/srep43630> (2017).
57. Hirst, S. A. & Maulik, S. On some Arthropod Remains from the Rhynie Chert (Old Red Sandstone). *Geological Magazine* **63**, 69–71 (1926).
58. Montañez, I. P. A. & Poulsen, C. J. The Late Paleozoic Ice Age: An Evolving Paradigm. *Annual Review of Earth and Planetary Sciences* **41**, 629–656 (2013).
59. Braslavsky, I. & Drori, R. LabVIEW-operated novel nanoliter osmometer for ice binding protein investigations. *J. Vis. Exp.*, e4189, <https://doi.org/10.3791/4189> (2013).
60. Marshall, C. J., Basu, K. & Davies, P. L. Ice-shell purification of ice-binding proteins. *Cryobiology*, <https://doi.org/10.1016/j.cryobiol.2016.03.009> (2016).
61. Tomalty, H. E., Graham, L. A., Eves, R., Gruneberg, A. K. & Davies, P. L. Laboratory-Scale Isolation of Insect Antifreeze Protein for Cryobiology. *Biomolecules* **9**, <https://doi.org/10.3390/biom9050180> (2019).
62. Clark, M. S. *et al.* Surviving the cold: molecular analyses of insect cryoprotective dehydration in the Arctic springtail *Megaphragma arctica* (Tullberg). *BMC Genomics* **10**, 328, <https://doi.org/10.1186/1471-2164-10-328> (2009).
63. Blankenberg, D. *et al.* A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. *Genome Res.* **17**, 960–964, <https://doi.org/10.1101/gr.5578007> (2007).
64. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gkz899> (2019).
65. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877, <https://doi.org/10.1101/gr.9.9.868> (1999).
66. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **47**(5 6), 1–32, <https://doi.org/10.1002/0471250953.bi0506s47> (2014).
67. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718, <https://doi.org/10.1002/jcc.20291> (2005).
68. DeLano, W. L. The PyMOL Molecular Graphics System. (2002).

Acknowledgements

This research was supported by the award of CIHR Foundation Grant FRN 148422 to PLD and Danish Research Council Grant DFF-4002-00384 to MH. PLD holds the Canada Research Chair in Protein Engineering. We are grateful to David McLeod of the Protein Function Discovery facility at Queen's University for the MALDI analysis of MaAFP.

Author contributions

L.A.G., P.L.D. and M.H. conceptualized the project. M.H. collected the animals. M.E.B. and L.A.G. acquired data. L.A.G., P.L.D. and M.H. interpreted the data. The manuscript was written by P.L.D., L.A.G. and M.H., and all authors contributed to its editing and reviewed the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-60060-z>.

Correspondence and requests for materials should be addressed to P.L.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020